

# IMGT/LIGM-DB Feature Table Definition Document

Version 13, Juin 2003

## CONTENTS

1. **INTRODUCTION**
2. **FORMAT EXAMPLE**
3. **FEATURE KEYS**
4. **FEATURE LOCATION**
5. **FEATURE QUALIFIERS**

## 1. INTRODUCTION

The feature table contains information about genes and gene products, as well as regions of biological significance reported in a sequence. It contains information on regions of the sequence that code for proteins and RNA molecules. It also enumerates differences between different reports of the same sequence and provides cross-references to other data collections, as described in more detail below.

The first two lines of the feature table in IMGT/LIGM-DB entries are feature header (FH) lines, specific to the EMBL flatfile format. The first one includes the column headers 'Key' and 'Location/Qualifier'. The second one is an empty spacer line.

Each feature consists of a feature key and a location (see below for details). If the location does not fit on the same line as the key, a continuation line may follow. If further information about the sequence is required, one or more additional lines containing feature qualifiers may follow.

Features appear on FT lines. The linetype code FT appears in columns 1-2 and columns 3-5 are blank. The feature key begins in column 6 and may be no more than 15 characters in length. The location begins in column 26. Feature qualifiers begin on subsequent FT lines at column 26. Location, qualifier, and continuation lines may extend from column 26 to 80. Each qualifier is added on a new line.

## 2. FORMAT EXAMPLE

An example of the feature table format is:

```

-----+-----+-----+-----+-----+-----+-----+
      10      20      30      40      50      60      70
      Key      Location/Qualifiers

L-PART1      1..28
V-GENE      1..222
              /cell_type="B cell"
              /note="NCBI gi: 483900"
              /partial
              /product="immunoglobulin kappa chain, V-region
              (SPK.4)"
              /tissue_type="Graves' thyroid"
    
```

```

-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      10      20      30      40      50      60      70

```

Thus, there are 4 types of feature table lines:

Line type	Content	#/entry	#/feature
-----	-----	-----	-----
Header	Column titles	1	N/A
Feature descriptor	Key and location	1 to many	1
Feature qualifiers	Qualifiers and values	N/A	0 to many
Continuation lines	Feature descriptor or qualifier continuation	0 to many	0 to many

The position of the data items within the feature descriptor line is as follows:

column position	data item
-----	-----
1-5	blank (may be used to improve readability, ie FT)
6-24	feature key
25	blank
26-80	location

Data on the qualifier and continuation lines begins in column position 26 (the first 25 columns contain blanks the first character is a '/' followed by the the qualifier discription). Qualifiers used here are the same as the EMBL qualifiers except for one exception the AA\_number qualifier.

The sections below provide a brief introduction to the new feature table format.

### 3. FEATURE KEYS

The first item on an FT line is the feature key. It starts in column 6 and can continue to column 24. The list of valid feature keys is shown below:

Label name	Definition
(DJ)-C-CLUSTER	genomic DNA in rearranged configuration including at least one D-J-GENE and one C-GENE
(DJ)-J-C-CLUSTER	genomic DNA in rearranged configuration including at least one D-J-GENE, one J-GENE and one C-GENE
(DJ)-J-CLUSTER	genomic DNA in rearranged configuration including at least one D-J-GENE, and one J-GENE
(VDJ)-C-CLUSTER	genomic DNA in rearranged configuration including at least one V-D-J-GENE and one C-GENE
(VDJ)-J-C-CLUSTER	genomic DNA in rearranged configuration including at least one V-D-J-GENE, one J-GENE and one C-GENE
(VDJ)-J-CLUSTER	genomic DNA in rearranged configuration including at least one V-D-J-GENE and one J-GENE
(VJ)-C-CLUSTER	genomic DNA in rearranged configuration including at least one V-J-GENE and one C-GENE
(VJ)-J-C-CLUSTER	genomic DNA in rearranged configuration including at least one V-J-GENE, one J-GENE and one C-GENE

(VJ)-J-CLUSTER	genomic DNA in rearranged configuration including at least one V-J-GENE and one J-GENE
1st-CYS	codon (3 nucleotides) for Cysteine in conserved position in FR1
2nd-CYS	codon (3 nucleotides) for Cysteine in conserved position in FR3
3'D-HEPTAMER	7 nucleotide recombination site like CACAGTG, part of a 3'D-RS
3'D-NONAMER	9 nucleotide recombination site like ACAAAAACC, part of a 3'D-RS
3'D-RS	recombination signal including the 3'D-HEPTAMER, 3'D-SPACER, and 3'D-NONAMER in 3' of the D-REGION of a D-GENE
3'D-SPACER	12 or 23 nucleotide spacer between the 3'D-HEPTAMER and 3'D-NONAMER of a 3'D-RS
3'UTR	3' untranslated sequence, EMBL feature Key signification
5'D-HEPTAMER	7 nucleotide recombination site like CACTGTG, part of a 5'D-RS
5'D-NONAMER	9 nucleotide recombination site like GGTTTTTGT, part of a 5'D-RS
5'D-RS	recombination signal including the 5'D-NONAMER, 5'D-SPACER and 5'D-HEPTAMER in 5' of the D-REGION of a D-GENE, or in 5' of the D-REGION of D-J-GENE
5'D-SPACER	12 or 23 nucleotide spacer between the 5'D-HEPTAMER and 5'D-NONAMER of a 5'D-RS
5'UTR	5' untranslated sequence, EMBL feature Key signification
ACCEPTOR-SPLICE	splicing site in 5' of coding region (naggn), with splicing occurring after g
C-CLUSTER	genomic DNA including more than one C-GENE
C-GENE	genomic DNA including C-REGION (and INTRONS if present) with 5' UTR and 3' UTR
C-LIKE-DOMAIN	coding region of non-IG and non-TR similar to an IG or TR C-DOMAIN
C-REGION	coding region of C-GENE or corresponding region in cDNA
C-SEQUENCE	cDNA including C-REGION (and INTRONS for unspliced cDNA) with 5' UTR and 3' UTR
CAAT_SIGNAL	'CAAT box' in eukaryotic promoters, EMBL Feature Key signification
CAP_SITE	m RNA cap site
CDR1	first complementarity determining region
CDR1-IMGT	first complementarity determining region according to the IMGT unique numbering
CDR2	second complementarity determining region
CDR2-IMGT	second complementarity determining region according to the IMGT unique numbering
CDR3	third complementarity determining region
CDR3-IMGT	third complementarity determining region according to the IMGT unique numbering
CH-S	3' end of CH3 or CH4 exon or independent exon which encodes the hydrophilic C-terminal end of soluble Ig, or corresponding region in cDNA
CH-SD	duplicated CH-S exon of IG heavy C-GENE (found in teleostei), or corresponding region in cDNA

CH-T	small terminal exon in truncated heavy chain transcript resulting of alternative splicing
CH-X	unusual exon of Ig heavy C-GENE, or corresponding coding region in cDNA
CH1	first exon of Ig heavy C-GENE, or corresponding coding region in cDNA
CH1D	duplicated CH1 exon of IG heavy C-GENE (found in teleostei), or corresponding region in cDNA
CH2	second exon of Ig heavy C-GENE (or part of the second exon when hinge sequence belongs to the same exon), or corresponding coding region in cDNA
CH2D	duplicated CH2 exon of IG heavy C-GENE (found in teleostei), or corresponding region in cDNA
CH3	third exon of Ig heavy C-GENE (including CH-S if present), or corresponding coding region in cDNA
CH3D	duplicated CH3 exon of IG heavy C-GENE (found in teleostei), or corresponding region in cDNA
CH4	fourth exon of Ig heavy C-GENE (including CH-S if present), or corresponding coding region in cDNA
CH4D	duplicated CH4 exon of IG heavy C-GENE (found in teleostei), or corresponding region in cDNA
CH5	fifth exon of Ig heavy C-GENE, or corresponding coding region in cDNA
CH6	sixth exon of Ig heavy C-GENE, or corresponding coding region in cDNA
CH7	seventh exon of Ig heavy C-GENE, or corresponding coding region in cDNA
CL	exon of Ig light C-GENE, or corresponding coding region in cDNA
CONFLICT	independent determinations differ, EMBL Feature Key signification
CONNECTING-REGION	coding region connecting the membrane proximal C-DOMAIN (or C-LIKE-DOMAIN) and the TRANSMEMBRANE-REGION
CONSERVED-TRP	codon (3 nucleotides) for Tryptophan in conserved position in FR2-IMGT
CYTOPLASMIC-REGION	coding intracytoplasmic region
D-(DJ)-C-CLUSTER	genomic DNA in rearranged configuration including at least one D-GENE, one D-J-GENE and one C-GENE
D-(DJ)-CLUSTER	genomic DNA in rearranged configuration including at least one D-GENE and one D-J-GENE
D-(DJ)-J-C-CLUSTER	genomic DNA in rearranged configuration including at least one D-GENE, one D-J-GENE, one J-GENE and one C-GENE
D-(DJ)-J-CLUSTER	genomic DNA in rearranged configuration including at least one D-GENE, one D-J-GENE, and one J-GENE
D-CLUSTER	genomic DNA in germline configuration including more than one D-GENE
D-GENE	germline genomic DNA including D-REGION with 5' UTR and 3' UTR,

	also designated as D-SEGMENT
D-J-C-CLUSTER	genomic DNA in germline configuration including at least one D-GENE, one J-GENE and one C-GENE
D-J-C-SEQUENCE	partially rearranged cDNA including D-, J- and C- REGION with 5'UTR and 3'UTR
D-J-CLUSTER	genomic DNA in germline configuration including at least one D-GENE and one J-GENE
D-J-GENE	partially rearranged genomic DNA including D-J-REGION with 5' UTR and 3' UTR, also designated as D-J-SEGMENT
D-J-REGION	coding region of D-J-GENE
D-J-SEQUENCE	partially rearranged cDNA including D- and J- REGION with 5'UTR and 3'UTR
D-REGION	coding region of D-GENE (plus 1 or 2 nucleotide(s) after the 5'D-HEPTAMER and/or before the 3'D-HEPTAMER, if present), or corresponding region in cDNA
D-SEQUENCE	germline cDNA including D-REGION with 5' UTR and 3' UTR
D1-REGION	coding region of the first D-GENE, when more than one D-GENE is involved in a JUNCTION, or corresponding coding region in cDNA
D2-REGION	coding region of the second D-GENE, when more than one D-GENE is involved in a JUNCTION, or corresponding coding region in cDNA
D3-REGION	coding region of the third D-GENE, when more than one D-GENE is involved in a JUNCTION, or corresponding coding region in cDNA
DECAMER	10 nucleotide regulation site or decanucleotide, includes OCTAMER, in the 5'UTR of a V-, V-D-, or V-D-J-GENE
DELETION	point out a deletion compared to other sequences
DONOR-SPLICE	splicing site in 3' of coding region (ngt), with splicing occurring before g
DUPLICATION	point out pattern duplication inside the sequence
ENHANCER	Cis-acting enhancer of promoter function, EMBL Feature Key signification
EX1	first exon of TcR C-GENE, or corresponding region in cDNA
EX2	second exon of TcR C-GENE, or corresponding region in cDNA
EX2A	exon 2A of TR C-GENE with exon 2 polymorphism by insertion/deletion or corresponding region in cDNA
EX2B	exon 2B of TR C-GENE with exon 2 polymorphism by insertion/deletion or corresponding region in cDNA
EX2C	exon 2C of TR C-GENE with exon 2 polymorphism by insertion/deletion or corresponding region in cDNA
EX2R	duplicated exon 2 of human TcR gamma C-GENE, or corresponding region in cDNA
EX2T	triplicated exon 2 of human TcR gamma C-GENE, or corresponding region in cDNA
EX3	third exon of TcR C-GENE, or corresponding region in cDNA
EX4	fourth exon of TcR C-GENE, or corresponding region in cDNA
EXON	exon of non Ig or non TcR genes, or corresponding coding region in cDNA

FR1	first framework
FR1-IMGT	first framework according to the IMGT unique numbering
FR2	second framework
FR2-IMGT	second framework according to the IMGT unique numbering
FR3	third framework
FR3-IMGT	third framework according to the IMGT unique numbering
FR4-IMGT	fourth framework according to the IMGT unique numbering
GENE	genomic DNA including EXONS and INTRONS with 5' UTR and 3' UTR and corresponding unspliced and spliced cDNAs for non-IG and non-TR genes
H	hinge exon of Ig heavy C-GENE, or corresponding region in cDNA
H1	first hinge exon of Ig heavy C-GENE, or corresponding region in cDNA
H2	second hinge exon of Ig heavy C-GENE, or corresponding region in cDNA
H3	third hinge exon of Ig heavy C-GENE, or corresponding region in cDNA
H4	fourth hinge exon of Ig heavy C-GENE, or corresponding region in cDNA
H5	fifth hinge exon of Ig heavy C-GENE, or corresponding region in cDNA
HEPTANUCLEOTIDE	7 nucleotide regulation site, like CTCATGC, in 5'UTR of a V-, V-D-, V-D-J-, or V-J-GENE
HINGE-REGION	coding region encoding the hinge in spliced cDNA
I-EXON	non coding exon located upstream of the switch, or corresponding region in cDNA
INDETERMINATION	point out an indetermination for a pattern
INIT-CODON	initiation codon ATG
INIT-CONS	consensus sequence upstream the INIT-CODON
INSERTION	point out an insertion of one or more nucleotides compared with old release of the sequence or with a similar sequence
INT-DONOR-SPLICE	alternative donor splice site located in a coding region
INTERNAL-HEPTAMER	internal 7 nucleotide recombination site in V-REGION
INTRON	transcribed region excised by mRNA splicing, EMBL Feature Key signification
J-C-CLUSTER	genomic DNA in germline configuration including at least one J-GENE and one C-GENE
J-C-INTRON	non coding region between the most 3' J-GENE and the following C-GENE, or corresponding sequence in unspliced cDNA
J-C-REGION	coding region including J- and C- REGION, in spliced cDNA
J-C-SEQUENCE	germline cDNA including J- and C-REGION (J-C-REGION in spliced cDNA, J-REGION, J-C-INTRON, and C-REGION in unspliced cDNA)
J-CLUSTER	genomic DNA in germline configuration including more than one J-GENE
J-GENE	germline genomic DNA including J-REGION with 5' UTR and 3' UTR,

	also designated as J-SEGMENT
J-HEPTAMER	7 nucleotide recombination site, like CACAGTG, part of a J-RS
J-NONAMER	9 nucleotide recombination site, like GGTTTTTGT, part of a J-RS
J-PHE	conserved phenylalanine in J-REGION of Ig light chain or TcR
J-REGION	coding region of J-GENE (plus 1 or 2 nucleotide(s) after J-HEPTAMER, if present) or corresponding region in cDNA
J-RS	recombination signal including J-HEPTAMER, J-SPACER and J-NONAMER in 5' of J-REGION of a J-GENE or J-SEQUENCE
J-SEQUENCE	germline cDNA including J-REGION with 5'UTR and 3'UTR
J-SPACER	12 or 23 nucleotide spacer between the J-NONAMER and the J-HEPTAMER of a J-RS
J-TRP	conserved tryptophan in J-REGION of Ig heavy chain
JUNCTION	coding region encompassing the V-J or V-D-J junction from 2nd CYS to the J-PHE or J-TRP of the J-REGION
L-INTRON-L	sequence including L-PART1, V-INTRON and L-PART2, in genomic DNA, or corresponding sequence in unspliced cDNA
L-PART1	exon encoding the first part of the leader peptide of a V-, V-D-, V-D-J- or V-J-GENE or corresponding region in unspliced cDNA
L-PART2	5' region of V-EXON encoding the second part of leader peptide of a V-, V-D-, V-D-J- or V-J-GENE or corresponding region in unspliced cDNA
L-REGION	coding region encoding the leader peptide in spliced cDNA
L-V-D-J-C-REGION	coding region including L-, V-, any D- and any N- REGION, J- and C- REGION, in cDNA
L-V-D-J-C-SEQUENCE	rearranged cDNA including L-REGION (or L-PART1 and L-PART2 for unspliced cDNA), V-, D-, J- and C-REGION with 5'UTR and 3'UTR
L-V-D-J-REGION	coding region including L-, V-, any D- and any N- REGION, and J- REGION, in cDNA
L-V-D-REGION	coding region including L-, V- and any D- and any N-REGION, in cDNA
L-V-D-SEQUENCE	partially rearranged cDNA including L-REGION (or L-PART1 and L-PART2 for unspliced cDNA), V- and D- REGION with 5'UTR and 3'UTR
L-V-J-C-REGION	coding region including L-, V-, J- and C- REGION, in cDNA
L-V-J-C-SEQUENCE	rearranged cDNA including L-REGION (or L-PART1 and L-PART2 for unspliced cDNA), V-, J- and C-REGION with 5'UTR and 3'UTR
L-V-J-REGION	coding region including L-, V-, and J- REGION, in cDNA
L-V-REGION	coding region including L- and V- REGION, in cDNA
L-V-SEQUENCE	germline cDNA including L-REGION (or L-PART1 and L-PART2 for unspliced cDNA) and V-REGION with 5' and 3'UTR
LINKER	short nucleotide sequence used to link 2 other nucleotide sequences
M	membrane exon of genomic C-GENE, or corresponding region in cDNA
M1	1st membrane exon of genomic C-GENE, or corresponding region in cDNA
M2	2nd membrane exon of genomic C-GENE, or corresponding region in cDNA

MISC_FEATURE	region of biological significance that cannot be described by other feature, EMBL Feature Key signification
MISC_RECOMB	Miscellaneous recombination feature, EMBL FeatureKey signification
MODIFICATION	shows a modification of the sequence or annotations compared to older release of the sequence or similar sequences
MUTATION	A mutation alters the sequence here, EMBL Feature Key signification
N-AND-D-J-REGION	coding region including N-AND-D- and J-REGION, in rearranged genomic DNA or corresponding region in cDNA
N-AND-D-REGION	coding region encompassing the N diversity sequences and coding region of D-GENE(s) in rearranged genomic DNA, or corresponding region in cDNA
N-GLYCOSYLATION-SITE	potential N glycosylation site encoded by the motif Asp-X-Ser/Thr where X is different from Pro
N-REGION	coding region encompassing the N diversity sequence
N1-REGION	coding region encompassing the first N diversity sequence, when more than one N-REGION is involved
N2-REGION	coding region encompassing the second N diversity sequence, when more than one N-REGION is involved
N3-REGION	coding region encompassing the third N diversity sequence, when more than one N-REGION is involved
N4-REGION	coding region encompassing the fourth N diversity sequence, when more than one N-REGION is involved
OCTAMER	8 nucleotide regulation site or octanucleotide, in the 5'UTR of a V-, V-D-, V-D-J-, or V-J-GENE
P-REGION	region encompassing the P sequence
PENTADECAMER	15 nucleotide regulation site or pentadecanucleotide, in the 5'UTR of a V-, V-D-, V-D-J-, or V-J-GENE
POLYA_SIGNAL	signal for cleavage & polyadenylation, EMBL Feature Key signification
POLYA_SITE	site at which polyadenine is added to mRNA, EMBL Feature Key signification
PRIMER_BIND	non-covalent primer binding site, EMBL Feature Key signification
PYR-RICH	rich pyrimidic bases regulation site, genomic gene
REPEAT_UNIT	one repeat unit of a repeat region, EMBL Feature Key signification
SILENCER	inhibitor signal for gene transcription, in genomic DNA
STERILE-TRANSCRIPT	unspliced or spliced cDNA corresponding either to a L-V-SEQUENCE, D-SEQUENCE, J-SEQUENCE or J-C-SEQUENCE in germline configuration, a L-V-D-SEQUENCE, D-J-SEQUENCE or D-J-C-SEQUENCE, or a C-SEQUENCE
STOP-CODON	codon which stops gene translation
SWITCH	switch sequence in the IGH locus
TATA_BOX	TATA signal in eukaryotic promoters
TRANSMEMBRANE-REGION	coding transmembrane region
UNSURE	authors are unsure about the sequence in this region, EMBL Feature Key signification



UTR	untranslated sequence
V-(DJ)-C-CLUSTER	genomic DNA in rearranged configuration including at least one V-GENE, one D-J-GENE and one C-GENE
V-(DJ)-CLUSTER	genomic DNA in rearranged configuration including at least one V-GENE and one D-J-GENE
V-(DJ)-J-C-CLUSTER	genomic DNA in rearranged configuration including at least one V-GENE, one D-J-GENE, one J-GENE and one C-GENE
V-(DJ)-J-CLUSTER	genomic DNA in rearranged configuration including at least one V-GENE, one D-J-GENE and one J-GENE
V-(VDJ)-C-CLUSTER	genomic DNA in rearranged configuration including at least one V-GENE, one V-D-J-GENE and one C-GENE
V-(VDJ)-CLUSTER	genomic DNA in rearranged configuration including at least one V-GENE and one V-D-J-GENE
V-(VDJ)-J-C-CLUSTER	genomic DNA in rearranged configuration including at least one V-GENE, one V-D-J-GENE, one J-GENE and one C-GENE
V-(VDJ)-J-CLUSTER	genomic DNA in rearranged configuration including at least one V-GENE, one V-D-J-GENE and one J-GENE
V-(VJ)-C-CLUSTER	genomic DNA in rearranged configuration including at least one V-GENE, one V-J-GENE and one C-GENE
V-(VJ)-CLUSTER	genomic DNA in rearranged configuration including at least one V-GENE and one V-J-GENE
V-(VJ)-J-C-CLUSTER	genomic DNA in rearranged configuration including at least one V-GENE, one V-J-GENE, one J-GENE and one C-GENE
V-(VJ)-J-CLUSTER	genomic DNA in rearranged configuration including at least one V-GENE, one V-J-GENE and one J-GENE
V-CLUSTER	genomic DNA in germline configuration including more than one V-GENE
V-D-(DJ)-C-CLUSTER	genomic DNA in rearranged configuration including at least one V-GENE, one D-GENE, one D-J-GENE and one C-GENE
V-D-(DJ)-CLUSTER	genomic DNA in rearranged configuration including at least one V-GENE, one D-GENE, one D-J-GENE
V-D-(DJ)-J-C-CLUSTER	genomic DNA in rearranged configuration including at least one V-GENE, one D-GENE, one D-J-GENE, one J-GENE and one C-GENE
V-D-(DJ)-J-CLUSTER	genomic DNA in rearranged configuration including at least one V-GENE, one D-GENE, one D-J-GENE and one J-GENE
V-D-EXON	partially rearranged genomic DNA including L-PART2, V-, any D- and N- REGION
V-D-GENE	partially rearranged genomic DNA including L-PART1, V-INTRON and V-D-EXON, with the 5'UTR and 3'UTR
V-D-J-C-CLUSTER	genomic DNA in germline configuration including at least one V-GENE, one D-GENE and one J-GENE and one C-GENE
V-D-J-C-REGION	coding region including V-, any D- and N- REGION, J- and C- REGION, in cDNA
V-D-J-CLUSTER	genomic DNA in germline configuration including at least one V-GENE, one D-GENE and one J-GENE
V-D-J-EXON	rearranged genomic DNA including L-PART2, V-, any D- and N-

	REGION, and J-REGION
V-D-J-GENE	rearranged genomic DNA including L-PART1, V-INTRON and V-D-J-EXON, with the 5'UTR and 3'UTR
V-D-J-REGION	coding region including V-, any D- and N-REGION, and J-REGION, in rearranged genomic DNA, or corresponding region in cDNA
V-D-REGION	coding region including V-, any D- and N- REGION, in rearranged genomic DNA or corresponding region in cDNA
V-EXON	germline genomic DNA including L-PART2 and V-REGION
V-GENE	germline genomic DNA including L-PART1, V-INTRON and V-EXON, with the 5'UTR and 3'UTR
V-HEPTAMER	7 nucleotide recombination site, like CACAGTG, part of V-RS
V-INTRON	non coding sequence between L-PART1 and V-EXON, in genomic DNA, or corresponding sequence in unspliced cDNA
V-J-C-CLUSTER	genomic DNA in germline configuration including at least one V-GENE, one J-GENE and one C-GENE
V-J-C-REGION	coding region including V-, J- and C- REGION, in cDNA
V-J-CLUSTER	genomic DNA in germline configuration including at least one V-GENE and one J-GENE
V-J-EXON	rearranged genomic DNA including L-PART2, V- and J- REGION
V-J-GENE	rearranged genomic DNA including L-PART1, V-INTRON and V-J-EXON, with the 5'UTR and 3'UTR
V-J-REGION	coding region including V- and J-REGION, in rearranged genomic DNA, or corresponding region in cDNA
V-LIKE-DOMAIN	coding region of non-IG and non-TR similar to an IG or TR V-DOMAIN
V-NONAMER	9 nucleotide recombination site, like ACAAAAACC, part of V-RS
V-REGION	coding region of V-GENE without the leader peptide (plus 1 or 2 nucleotide(s) before the V-HEPTAMER, if present), or corresponding region in cDNA
V-RS	recombination signal including V-HEPTAMER, V-SPACER and V-NONAMER in 3' of V-REGION of a V-GENE or V-SEQUENCE
V-SPACER	12 or 23 nucleotide spacer between the V-HEPTAMER and the V-NONAMER of a V-RS
VARIATION	a related population contains stable mutations, EMBL Feature Key signification
scFv	defines two immunoglobulin (or by extension T cell receptor) V-DOMAINS covalently linked by a short linker peptide in vitro

#### 4. FEATURE LOCATION

The second item on the FT line designates the location of the feature in the sequence. The location begins at column 26. Several conventions are used to indicate sequence location.

Base numbers in locations refer to the numbering in the entry, which is not necessarily the same as the numbering scheme used in the original report. The first base in the presented sequence is numbered base 1. Sequences are presented in the 5' to 3' direction.

A location can be one of the following:

- o A single base.
- o A contiguous span of bases.

A contiguous span of bases is indicated by the number of the first and last bases in the range separated by two periods (e.g., 23..79). Starting and ending positions can be indicated by base number.

## 5. FEATURE QUALIFIERS

Qualifiers provide additional information about features. They take the form of a slash (/) followed by a qualifier name and, if applicable, an equal sign (=) and a qualifier value. Feature qualifiers begin at column 26.

Qualifiers convey many types of information. Their values can, therefore, take several forms:

- o Free text.
- o Controlled vocabulary or enumerated values.
- o Citations or reference numbers.
- o Sequences.
- o Feature labels.

Text qualifier values are enclosed in double quotation marks. The text can consist of any printable characters (ASCII values 32-126 decimal). If the text string includes double quotation marks, each double quotation mark must be escaped by placing a double quotation mark in front of it (e.g., /note="This is an example of ""escaped"" quotation marks").

Citation or reference numbers for an entry are enclosed in square brackets ([ ]) to distinguish them from other numbers.

A literal sequence of bases (e.g., "atgcatt") is enclosed in quotation marks. Literal sequences are distinguished from free text by context. Qualifiers that take free text as their values do not take literal sequences, and vice versa.

The /label= qualifier takes a feature label as its qualifier. Although feature labels are optional, they allow unambiguous references to features. The feature label identifies a feature within an entry; when combined with the accession number and the name of the data bank from which it came, it is a unique tag for that feature.

The following is a list of valid feature qualifiers:

Qualifier	Description
allele	Name of the allele for the a given gene
allotype	polymorphic extracellular marker detected by serological methods and present in different individuals of the same species
AA_IMGT	Amino Acid numerotation in the sequence according to IMGT
AA_number	Amino Acid numerotation in the sequence

cell_line	Cell line from which the sequence was obtained
cell_type	Cell type from which the sequence was obtained
chromosome	Chromosome (e.g. Chromosome number) from which the sequence was obtained
citation	Reference to a citation listed in the entry reference field
clone	Clone from which the sequence was obtained
clone_lib	Clone library from which the sequence was obtained
codon_start	Indicates the offset at which the first complete codon of a coding feature can be found, relative to the first base of that feature
cons_splice	Differentiates between intron splice sites that conform to the 5'-GT ... AG-3' splice site consensus
country	Country of origin for DNA sample, intended for epidemiological or population studies
CDR_length	Number of Amino Acids in CDR1-IMGT, CDR2-IMGT, CDR3-IMGT, separated by dots, and shown in brackets. X is used for partial or absent CDR
db_xref	Database cross-reference: pointer to related information in another database
dev_stage	If the sequence was obtained from an organism in a specific developmental stage, it is specified with this qualifier
evidence	Value indicating the nature of supporting evidence, distinguishing between experimentally determined and theoretically derived data
function	Function attributed to a sequence
gdb_xref	Genome Databank unique ID cross reference qualifier
gene	Symbol of the gene corresponding to a sequence region
gene_alias	other gene name in the literature
germline	Denotes that the sequence is from immunoglobulin or T cell receptor unarranged DNA or RNA
germline_frame	Translation arbitrarily shown in the germline reading frame, for J-REGION (and C-REGION in cDNA) of unproductive (genomic or cDNA) rearranged sequences
haplotype	Haplotype of the organism from which the sequence was obtained
insertion_seq	Insertion sequence element from which the sequence was obtained
in_frame	No frameshift in the JUNCTION
isolate	Individual isolate from which the sequence was obtained
IMGT_BAC_clone	Name of the BAC clone from which the sequence is derived
IMGT_cell_line	Name of the cell line from which the sequence is derived
IMGT_cosmid_clone	Name of the cosmid clone from which the sequence is derived
IMGT_MAC_clone	Name of the MAC clone from which the sequence is derived
IMGT_note	comment added by the LIGM curators to the IMGT feature
IMGT_phage_clone	Name of the phage clone from which the sequence is derived
IMGT_plasmid_clone	Name of the plasmid clone from which the sequence is derived
IMGT_YAC_clone	Name of the YAC clone from which the sequence is derived
label	A label used to permanently identify a feature
lab_host	Laboratory host used to propagate the organism from which the sequence was obtained

map	Genomic map position of feature
nomgen	Name of the gene corresponding to a sequence region
note	Any comment or additional information
number	A number to indicate the order of genetic elements (e.g., exons or introns) in the 5' to 3' direction
organism	The scientific name of the organism that provided the sequenced genetic material
out_of_frame	frameshift in the JUNCTION
partial	Differentiates between complete regions and partial ones
product	Name of a product encoded by the sequence
protein_id	Protein Identifier, issued by International collaborators. This qualifier consists of a stable ID portion (3+5 format with 3 position letters and 5 numbers) plus a version number after the decimal point.
pseudo	Indicates that this feature is a non-functional version of the element named by the feature key
putative_limit	Refers to uncertain limit(s) of a subregion
PCR_conditions	Description of reaction conditions and components for PCR
rearranged	Denotes that the sequence is from immunoglobulin or T cell receptor rearranged DNA or RNA
replace	indicates that the sequence identified by a feature's intervals is replaced by the sequence shown in "text"
rpt_family	Type of repeated sequence; Alu or Kpn, for example
rpt_type	Organization of repeated sequence
rpt_unit	Identity of repeat unit that constitutes a repeat region
sequenced_mol	Molecule from which the sequence was obtained
sex	Sex of organism from which the sequence was obtained
specificity	Specificity of an immunoglobulin or T cell receptor chain
specific_host	Natural host from which the sequence was obtained
standard_name	Accepted standard name for this feature
strain	Strain from which the sequence was obtained
sub_clone	Sub-clone from which the sequence was obtained
sub_species	Sub-species name of organism from which the sequence was obtained
sub_strain	Sub-strain from which the sequence was obtained
tissue_lib	Tissue library from which the sequence was obtained
tissue_type	Tissue type from which sequence was obtained
translation	Automatically generated one-letter abbreviated amino acid sequence of the coding regions
transl_except	Translational exception: single codon the translation of which does not conform to genetic code defined by Organism and /codon

This manual and the database it accompanies may be copied and redistributed freely, without

Last modified: Juin 2003

Software material and data coming from IMGT server may be used for academic research only, provided that it is referred to IMGT, and cited as "IMGT, the international ImMunoGeneTics database <http://imgt.cines.fr:8104> (Initiator and coordinator: Marie-Paule Lefranc, Montpellier, France)." References to cite: Lefranc, M.-P. et al., *Nucleic Acids Research*, 27, 209-212 (1999); Ruiz, M. et al., *Nucleic Acids Research*, 28, 219-221 (2000) Lefranc, M.-P., *Nucleic Acids Research*, 29, 207-209 (2001), *Nucleic Acids Res.*, 31, 370-310 (2003) [Full text](#).

For any other use please contact Marie-Paule Lefranc [lefranc@ligm.igh.cnrs.fr](mailto:lefranc@ligm.igh.cnrs.fr).

---

IMGT initiator and coordinator: Marie-Paule Lefranc ([lefranc@ligm.igh.cnrs.fr](mailto:lefranc@ligm.igh.cnrs.fr))

Bioinformatics manager: Véronique Giudicelli ([giudi@ligm.igh.cnrs.fr](mailto:giudi@ligm.igh.cnrs.fr))

Computer manager: Denys Chaume ([Denys.Chaume@igh.cnrs.fr](mailto:Denys.Chaume@igh.cnrs.fr))

Interface design: Chantal Ginestoux ([chantal@ligm.igh.cnrs.fr](mailto:chantal@ligm.igh.cnrs.fr))

© Copyright 1995-2003 IMGT, the international ImMunoGeneTics database